# The case for risk-based premiums in public health insurance

Zweifel, Peter ; Breuer, Michael

Abstract: Uniform, risk-independent insurance premiums are accepted as part of 'managed competition' in health care. However, they are not compatible with optimality of health insurance contracts in the presence of both ex ante and ex post moral hazard. They have adverse effects on insurer behaviour even if risk adjustment is taken into account. Risk-based premiums combined with means-tested, tax-financed transfers are advocated as an alternative.

DOI: https://doi.org/10.1017/S1744133105001064

*Health Economics, Policy and Law* (2006), **1**: 171–188 Printed in the United Kingdom
 doi:10.1017/S1744133105001064

## Debate

# The case for risk-based premiums in public health insurance

PETER ZWEIFEL* AND MICHAEL BREUER
*Socioeconomic Institute of the University of Zurich, Switzerland*

**Abstract:** Uniform, risk-independent insurance premiums are accepted as part of 'managed competition' in health care. However, they are not compatible with optimality of health insurance contracts in the presence of both ex ante and ex post moral hazard. They have adverse effects on insurer behaviour even if risk adjustment is taken into account. Risk-based premiums combined with means-tested, tax-financed transfers are advocated as an alternative.

### 1. Introduction

In the course of the past decade, several countries have turned to competition between public health insurers in the hope of spurring innovation that ultimately would contribute to controlling the surge of health care expenditure. The concept is that of 'managed competition', implying that health insurers continue to be subject to many regulations (Enthoven, 1986). Among these regulations, uniform, risk-independent premiums are accepted as a given, with e.g. the imposition of community rating at the level of employees of a firm in the United States, a uniform contribution rate in terms of labour income to be charged by a given insurer in Germany and the Netherlands, and uniform absolute premiums to be charged by a given insurer in Switzerland. The common feature of these (and other) variants is that differences in expected loss must not be reflected in contributions. The present contribution purports to show that uniformity of health insurance premiums entails a considerable loss in terms of efficiency, while not necessarily serving accepted distributional goals. Open enrolment is assumed throughout because otherwise competition would be stifled; however, open enrolment makes self-selection of risks possible, which induces additional regulation. It will be shown that this regulation, while failing to prevent fully risk selection, causes efficiency loss. The contribution also contains a policy proposal for an alternative solution designed to reach

* Corresponding author: Socioeconomic Institute of the University of Zurich, Hottingerstrasse 10, CH-8032 Zurich, Switzerland. Tel: +41 1 634 37 20; Email: pzweifel@soi.unizh.ch

distributional objectives, while reaping the efficiency gains to be had by permitting premiums to reflect risk.

The next section is devoted to a review of the properties of optimal health insurance contracts in the presence of moral hazard. The distinction between *ex ante* and *ex post* moral hazard is relevant for the definition of the term 'risk' in calculating the premium. In both cases, however, theory indicates that premiums must reflect risk to achieve efficiency. Conversely, uniformity of premiums causes efficiency losses originating from non-optimal behaviour of the insured. Next, Section 3 focuses on the consequences of imposing uniform premiums for insurer behaviour. The presence of a risk adjustment scheme designed to counteract the increased incentive for risk selection is assumed throughout. The finding is that even refined schemes cannot meet this objective. In Section 4, the same analysis is performed with regard to product innovation to conclude that risk adjustment schemes of the type presently known inherently punish innovative effort on the part of insurers. Since the several failures identified are ultimately due to the regulation of health insurance premiums, Section 5 sketches an alternative. It consists of risk-based premiums, complemented by a tax-financed transfer mechanism. Through means-tested premium subsidies, the public purse takes over the redistribution function from health insurers. A rough calculation for the case of Germany indicates that the budgetary consequences become rather small, once the present shortfall in tax revenue caused by shared employer and employee contributions for health insurance are taken into account. The final section concludes.

## 2. Incompatibility of optimal health insurance contracts with uniform premiums

In this section, the properties of optimal insurance contracts are reviewed with special reference to health insurance. The objective is to obtain a benchmark against which a contract imposing uniform, risk-independent premiums can be compared and to gauge the severity of efficiency loss imparted. Thus, focus is on efficiency because uniform premiums need not guarantee equity. Indeed, they result in a cross-subsidization of high-risk, high-income individuals by low-risk, low-income individuals. This can result in counter-productive effects. For example, a healthy young blue-collar worker subsidizes a wealthy older manager who is a heavy user of medical services. Equity considerations seem to call for redistribution from everyone else to the doubly disadvantaged, viz. the high-risk, low-income individuals. This requirement is taken up in Section 5, on risk-based premiums and transfers as an alternative.

For the optimality of an insurance contract, the main difference is whether there is moral hazard or not. Without moral hazard, optimal contracts could be designed exclusively to prevent adverse selection. Since this involves insurer behaviour at least to some extent, consideration of this case is deferred to Section 3. With moral hazard present, the objective of an insurance contract

becomes to induce overall optimal behaviour of the insured *ex ante* (in the guise of preventive effort) as well as *ex post* (in the guise of HCE in the event of illness). This distinction also points to the definition of the term 'risk' when talking about risk-based premiums. In the context of *ex ante* moral hazard, this is the probability of illness ($\pi$) resulting from preventive effort ($V$). In the case of *ex post* moral hazard, it is the size of the loss ($L$) given that illness has occurred.

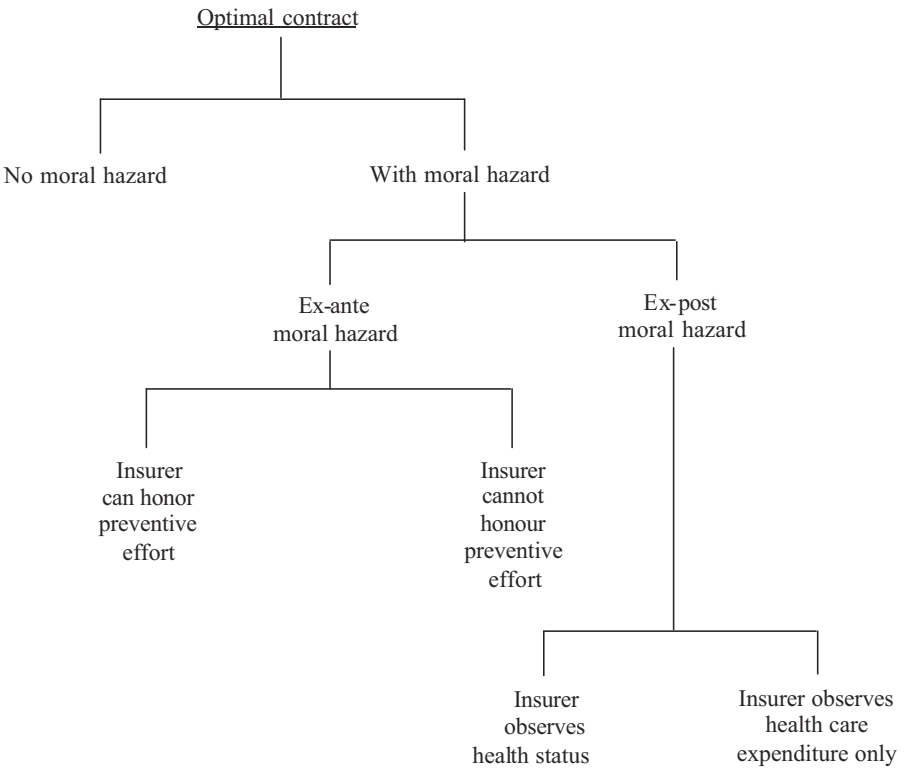
### *2.1 Optimal contracts in the presence of ex ante moral hazard*

*Ex ante* moral hazard would not be a problem if the premium could be made to depend on the probability of illness as it results from preventive effort. The insurer would simply calculate the premium according to the observable probability of illness $\pi$, which is the result of preventive effort $V$. The premium reduction for preventive effort would become the marginal return to the insured, who would balance it against the marginal cost of prevention (Ehrlich and Becker, 1972). Some distortions cannot be excluded even in this ideal situation, e.g. because a loading magnifies the marginal return in terms of premium saved or because the insurer does not find it worthwhile to obtain a precise estimate of $\pi$ in view of cost. Still, this first-best solution can be attained only if premiums are allowed to vary with preventive effort and, therefore, with the resulting probability $\pi$.

Of course, the far more relevant case is the one of *ex ante* moral hazard, with the insurer unable to honour preventive effort (see Figure 1 again). Here, the question arises as to whether the insurer can devise a contract that induces at least some preventive effort (in keeping with the incentive compatibility constraint), while getting the insured to buy it (participation constraint). For simplicity, the argument is couched in an expected utility framework. This does not constitute a severe loss of generality in view of the robustness of comparative static results with respect to leading non-expected utility alternatives (Machina, 1995; Gollier, 2000). Moreover, direct utility effects of health are neglected unless otherwise stated. Again, the loss of generality is limited. The marginal returns to prevention would have to be augmented by the financial equivalent of improved health. Therefore, it is conceivable that some preventive effort is forthcoming without any contractual incentives. Whether this would be the optimal amount is uncertain, however, and contractual incentives may still be important for reaching the optimal level.

What the simplified model predicts is that: given full insurance coverage, the optimal level of prevention is zero. Conversely, it takes less than full coverage to induce a positive amount of preventive effort. In order to satisfy the participation constraint, the insurer must reduce the premium accordingly so that the expected utility of the insured can be higher than the alternative with full coverage with its associated premium in case of zero preventive effort. Individuals who can count on a high marginal effectiveness of their preventive effort will

**Figure 1.** Optimal health insurance contracts




Source: Zweifel and Breyer (1997), Ch. 6.1

see their participation constraint satisfied at a rather small reduction in premium. These considerations again result in the requirement that the optimal premium must reflect $\pi$ as it results from preventive effort. Since the amount of cost sharing for optimally limiting *ex ante* moral hazard depends also on individual parameters (specifically, marginal effectiveness and opportunity cost of prevention), premiums and premium differentiation according to cost sharing cannot be uniform for optimality.

At this point, an element of regulation should be mentioned that must be retained in any 'managed competition' concept. As argued by Pauly (1974), the insured may try to combine zero preventive effort ($V = 0$) with paying a premium calculated for $V > 0$ by signing more than one contract, each specifying some cost sharing and hence a premium for $V > 0$, but with the joint effect of granting full coverage, making $V = 0$ optimal. To prevent this accumulation of coverage, insurers should be allowed to share information with regard to total coverage of their clients. If the public purse contributes to the payment of the premium in any way (through hidden tax subsidies or explicit subsidies

as proposed in Section 5 below), there is an interest in requiring a minimum degree of cost sharing. In this event, regulation would have even to identify a leading insurer that fixes the overall amount of cost sharing, with supplementary insurers being prohibited from letting total coverage exceed that threshold.

**Conclusion 1:** In the presence of *ex ante* moral hazard, incentive compatibility generally calls for partial insurance coverage, and the participation constraint, for a lowered premium. Premiums thus optimally are risk-based.

### *2.2 Optimal contracts in the presence of ex post moral hazard*

Once some illness has materialized, again an issue of observability arises. If health status is observable or it is made observable by a physician as an impartial observer, then indemnity insurance is optimal (Zeckhauser, 1970). The indemnity is fixed in a way as to make the final health status identical across illnesses. In this way, the marginal utility of wealth can be made equal across all states of health. Since the marginal cost of medical care is fully borne by the insured beyond the indemnity, the choice of intensity of treatment can be left to the insured. This serves to neutralize fully the propensity of moral hazard the insured might have. Still, this uniformity of benefits does not imply uniformity of premiums because probabilities of different illness states may differ between individuals. In a competitive health insurance market, premiums that make optimal indemnity contracts viable must therefore be risk-based again.

The far more relevant case of *ex post* moral hazard is of course when health status is unobservable. What remains observable is health care expenditure (HCE), and the optimizing instrument in the hands of the insurer is once more the degree of cost sharing. As shown by Zeckhauser (1970) and more recently Blomqvist (1997), a cost-sharing rule that is non-linear in expenditure turns out to be generally optimal. For example, Buchanan *et al*. (1991) simulate optimal insurance policies to arrive at a $200 deductible and a 25% coinsurance rate, whereas Blomqvist (1997) comes up with a cost sharing of 27% at roughly $1,000 and declining to 5% at roughly $30,000. The parameters of the rule depend on individual characteristics, implying that uniform premiums once more cannot be optimal.

If the analysis is restricted to a linear cost-sharing rule involving the single coinsurance parameter $c$, the value of $c$ optimally depends on the elasticity of the demand for medical care with regard to $c$, which amounts to a price elasticity (Zweifel and Breyer, 1997, Chapter 6.5). The more marked this price elasticity, the more important the *ex post* moral hazard effect, which needs to be counteracted by cost sharing in the insurance contract. In a generalized analysis with many health states, their probabilities along with the corresponding

marginal utilities of wealth enter as additional parameters. Finally, the generosity of sick leave pay affects *ex post* moral hazard effects on the demand for health care (Zweifel and Manning, 2000), which suggests that optimal health insurance depends on additional parameters. Uniform premiums are not able to accommodate these individual differences in price elasticities, probabilities, and parameters from other insurance contracts.

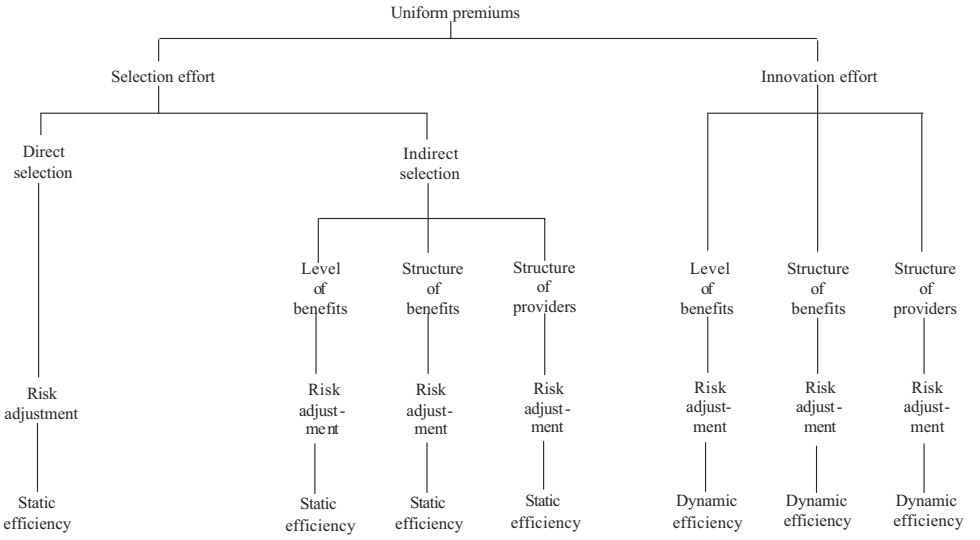
**Conclusion 2:** In the presence of *ex post* moral hazard, cost sharing in response to individual parameters such as price elasticities of demand for medical care is optimal, resulting in risk-based premium reductions and hence premiums.

Conclusions 1 and 2 are in terms of conventional health insurance, and we may wonder whether they are relevant for managed care. In fact, a managed-care insurer must take into account *ex ante* moral hazard in much the same way as a conventional one because $\partial\pi/\partial V$, the relationship between preventive effort and probability of illness, remains fully under the influence of the insured. After all, physicians do not seem to have much effect on the patients' probability of falling ill, whether working for a managed-care or conventional insurance organization. In the case of *ex post* moral hazard, the issue becomes the extent to which the insured retain some degree of influence on HCE. In the extreme case of no cost sharing at all, the managed-care insurer must rely on physicians as rationing agents. The insured then select a contract not according to cost-sharing parameters as in conventional insurance but according to the importance of insurer considerations in medical decision making, which is maximum in the staff HMO model and minimum for Independent Practice Associations (Glied, 2000). As long as physicians practise different styles when dealing with different types of managed-care patients, an *ex post* moral hazard effect exists, which can be counteracted by charging a higher premium for granting access to physicians that are more accommodating than others. Therefore, uniform premiums again are not compatible with optimality, although the incompatibility is less marked than in conventional medical care.

## 3. Uniform premiums, risk adjustment, and health insurer behaviour

This section is devoted to an analysis of health insurer behaviour under the constraint that premiums cannot be set according to risk. This means that, for breaking even, insurers must keep total expenditure per insured below the uniform premium that the market will bear for their differentiated product. To achieve this goal, they have two instruments at their disposal. One is risk selection effort, the other, cost-reducing innovation effort, which will be analyzed in Section 4.

Figure 2. Effects of uniform premium regulation on insurer behaviour




### *3.1 Risk adjustment as induced regulation*

If selection and cost-reducing innovation are two instruments available to health insurers, the factors affecting their balance are of interest. The most important factor is premium regulation (see Figure 2). Indeed, a necessary condition for selection to be profitable is that the contribution margins of high and low risks differ (Pauly, 1984). Uniform premiums (and indeed all forms of premium that are not risked-based) by definition cause this divergence because an individual with high expected future HCE cannot be made to contribute a higher premium. Quite frequently, expected HCE even exceeds the premium received. Simply to break even, the insurer must have a sufficient number of individuals enrolled, whose expected HCE is below the uniform premium. This creates an incentive to undertake selection effort.

Thus, it is the premium regulation (especially in the guise of uniform premiums) that tilts the balance in favour of selection and against cost-reducing innovation. Additional regulation becomes necessary to redress the balance. For an extended list of measures, see Van de Ven and Ellis (2000); however, risk adjustment occupies the prominent place among them. Risk adjustment is a mechanism that equalizes the difference between premiums received and expected HCE across risk groups. A risk adjustment scheme receives payments from insurers enrolling with a below-average share of high risks and uses these funds to disburse payments to insurers enrolling with an above-average share of high risks. In all, risk adjustment can bee seen as a secondary regulation induced by the primary regulation of uniform premiums, designed to neutralize its undesired side effects.

In the following, it will be argued that risk adjustment necessarily is imperfect, thus failing to fully counteract the distortion of incentives caused by uniform premiums. One such failure is of a general type and can be shown at this point.

In order to keep insurers' incentives for cost control in tact, risk adjustment needs to be prospective. This means that risk adjustment should be based on predicted future rather than actual HCE. However, this burdens insurers with unexpected deviations in HCE from their expected value. To the extent that these insurers are mutuals rather than companies quoted on the stock exchange, their owners have minimum possibility to get their insurance managers to act in a risk-neutral manner. Therefore, prospective risk adjustment is resisted by the industry and in several countries given up in favour of retrospective risk adjustment (Van de Ven and Ellis, 2000). Other failures are more specific to direct and indirect selection, to be expounded below.

### *3.2 Problems of risk adjustment and direct selection*

Direct selection occurs if the insurer observes certain characteristics of an individual that allow a forecast of future HCE (see Figure 2). Notably, age, sex, and pre-existing health conditions can be used to screen high risks, who are not enrolled at all if possible. For example, application forms are not sent and not processed or processed with exceedingly long delays in the hope that the applicant loses patience. Alternatively, the representative of the insurer may even point out a competitor as being more suitable. Of course, this violates the insurers' obligation to contract; however, such obligation is often costly to enforce. Moreover, obligation to contract does not preclude indirect selection (see Section 3.3). The alternative is to put in place a risk adjustment scheme that neutralizes the differences in expected HCE that are associated with a characteristic. This means that the regulator must be able to relate HCE to the relevant characteristics (called risk adjusters).

Existing research shows that on the basis of publicly observable risk adjusters, such as age, sex, and place of residence, only 3–4% of the variance in annual individual HCE can be predicted (Van de Ven and Van Vliet, 1992; Beck and Zweifel, 1998). This leaves a great deal of scope for the use of private information that insurers will want to exploit in order to improve their own predictions. Therefore, risk adjustment cannot compensate fully for differences in HCE due to risk characteristics (and other factors beyond the control of health insurers, such as regional differences in input prices). Including previous HCE in the set of risk adjusters does increase the explained share of the HCE variance, but at the price of making risk adjustment partially retrospective. Moreover, some 80% of the HCE variance still remains unexplained (Van Vliet, 1992).

**Conclusion 3:** Even partially retrospective risk adjustment falls far short of neutralizing health insurers' incentives for direct risk selection under uniform premium regulation.

As a consequence, risk adjustment fails to re-establish the conditions for insurer behaviour that is compatible with static efficiency, which would be characterized by an absence of risk selection effort.

In all, the incentive for indirect risk selection fails to be fully neutralized.

### *3.3 Problems of risk adjustment and indirect selection*

Indirect selection occurs if insurers rely on unobserved differences in risk types. All they have to know is what risk types exist in the population and how risk types are related in general ways with preferences for certain contract types. Indirect selection can be operated through the choice of (a) benefit level, (b) benefit structure, and (c) provider structure.

(a) **Indirect selection through the benefit level:** In the simplest case of two risk types differing only in terms of their illness probability $\pi$, this amounts to offering a contract with high coverage and high premium to high risks and low coverage and low premium to low risks. Of course, this is nothing but the 'separating contracts' solution to the adverse selection problem in an unregulated private insurance market (Rothschild and Stiglitz, 1976). In that model, the alternative is a pooling contract that has been shown not to be sustainable, unless there are few competitors in the market. In that case, an insurer having a two-period planning horizon will realize that challenging a competitor's pooling contract will backfire, because the high risks, having lost coverage, will end up in its own insured population in the second period (Wilson, 1977). Given premium regulation, the pooling contract is legally enforced. The financial sanction associated with deviating is the additional payment to the risk adjustment scheme. In expected value, it must equal the returns to be reaped from letting indirect selection work. In the case of benefit levels, these returns depend on the following factors:
   (1) The difference in expected HCE caused by the characteristic that contributes to the indirect selection process;
   (2) The probability that the difference in the characteristic level still exists during the future periods making up the planning horizon of the insurer;
   (3) The discounting factor used by the insurer;
   (4) The cost of selection effort, i.e. of developing and launching a set of contracts with benefit levels that effectively trigger the desired indirect selection process, plus the financial consequences due to risk adjustment.

An optimal risk adjustment scheme would have to set payments into and receipts from the scheme in a way that their expected value results in a balance of zero of these four factors for each health insurer. However, present-day schemes fall far short of taking due account of these factors:

(1′) Estimates of expected HCE in relation to observable characteristics are biased. Since current risk adjustment formulae are limited to a few observable indicators, which in a regression jointly explain few percent of the variance of annual HCE, unobserved characteristics are important. These are left-out explanatory variables that generally result in biased coefficients of the

included variables, and hence of estimated differences in HCE associated with age, e.g. diagnostic information. If such information can be made available legally, it does not fully solve the problem because there are still unobserved variables. Specifically, in section 2.2 it was found that the price elasticity of the demand for medical care in different health states influences HCE.

(2′) A low risk may no longer be a low risk two or three years in the future, and vice versa for the high risks. Indeed, Beck and Zweifel (1998) found evidence of considerable transition probabilities. Moreover, these transition probabilities reflect *ex ante* moral hazard, which is affected by the structuring of contracts.

(3′) The discount factor to be applied should in principle be the risk-adjusted rate of return on the capital market; however, since health insurers typically are not quoted on the stock market, the discount factor becomes private information.

(4′) The cost of contract development and launch is insurer-specific and constitutes private information. In addition, the regulator would have to know the starting point from where risk adjustment payments are to be calculated. On the one hand, observed values of the two types of HCE may still reflect a pooling equilibrium, where the two types are free to choose according to their risk types. On the other hand, it may already correspond to the separating equilibrium, with limits on coverage in the contract for low risks. In the latter case, payments are insufficient to neutralize fully the incentive to make indirect selection efforts.

**Conclusion 4:** Risk adjustment is heavily imperfect with regard to indirect selection through benefit levels.

An alternative to risk adjustment would be to also require uniformity of benefit levels and hence cost sharing. However, this would severely restrict insurers' possibilities to control moral hazard.

(b) **Indirect selection through the benefit structure:** Here, the two risk types can be defined, e.g. by their different probabilities of consuming medical care for a chronic condition, in addition to medical care for an acute one. The imposition of a uniform premium causes the budget constraint for the high risks to be shifted inward in terms of the two types of medical care and outward for the low risks. As shown by Glazer and McGuire (2000), the two types (being characterized by different values of $\pi$ for a chronic condition) have different preference structures, a fact that can be used to initiate indirect selection processes. The returns to indirect selection through the benefit structure depend on the following factors:

(1) The difference in expected HCE caused by the high risks' preference for the more costly chronic care;

(2) The discounting factor used by the insurer;

(3) The cost of selection effort, i.e. of developing and launching contracts that are stingy enough in terms of chronic care to keep high risks out while attracting low risks, plus the financial consequences due to risk adjustment.

Contrary to case (a), transition probabilities do not enter here because of the assumed chronicity. Again, the regulator operating a risk adjustment scheme faces the following problems.

(1′) Expected HCE of chronic patients is influenced by moral hazard effects to a particularly high degree because of their higher level of information about more or less costly treatment alternatives. This constitutes a specific left-out risk adjuster that causes bias in estimated HCE differentials.
(2′) The discounting factor is private information, for the reason given above in Conclusion 4.
(3′) The rationing of chronic medical care introduces the same ambiguity as noted in (a), point (4′).

An alternative to risk adjustment is to simply impose uniformity with regard to the structure of insurance benefits. However, this would run entirely counter to the idea of competition in health insurance, which promises an improved matching of benefits provided and preferences of insureds. Another solution is the institutional separation of the types of benefits (Van de Ven and Ellis, 2000). However, this may entail considerable loss of economies of scope (in the packaging of different benefits) and advantages from substitution (in the actual delivery of care).

**Conclusion 5:** Risk adjustment is also heavily imperfect with regard to indirect selection through the benefit structure.

(c) **Indirect selection through the provider structure:** This is conceptually very similar to selecting through the benefit structure. High risks are insured that are particularly likely to call on the services of higher-cost providers [rather than providers of chronic care, as in (b) above]. The imposition of a uniform premium causes the budget constraint for the high risks to be shifted inward in terms of the services rendered by the two provider types and outward for the low risks (Glazer and McGuire, 2000). Here, the differences between the budget constraints characterizing the different types of insured are accentuated because the low-cost provider may make special efforts to contain moral hazard on the part of the insured. The returns to indirect selection thus depend on the following factors:
(1) The difference in expected cost associated with the type of provider contracted with;
(2) The difference in probability of insured calling on high-cost rather than low-cost providers;
(3) The probability that the contractual relationship with a given type of provider still exists during the future periods making up the planning horizon of the insurer;
(4) The discounting factor used by the insurer;
(5) The cost of selection effort, i.e. of negotiating contracts with particular types of providers and launching the corresponding insurance policies.

Again, present-day risk adjustment schemes are not attuned to these factors.

(1′) Expected HCE differs also because insurers have different ways to control moral hazard effects among service providers. These effects would have to be netted out lest insurers be punished for introducing new, efficiency-enhancing contracts.
(2′) The reduced probability of visiting high-cost providers may again be the result of efforts on the part of the insurer. Failure by the regulator to take this into account creates the risk of punishing cost-reducing innovation.
(3′) A given group of service providers may not be the same even after a single year, mainly because of failed contract renewal. For *ex ante* risk adjustment, these changes need to be accounted for.
(4′) The discounting factor continues to constitute private information.
(5′) The cost of negotiation with a particular group of providers must remain private information because it importantly influences the insurer's bargaining situation.

**Conclusion 6:** When it comes to indirect risk selection through the provider structure, risk adjustment is capable of neutralizing incentives to an even lesser degree than in the cases of selection through the benefit level (a) and the benefit structure (b).

Moreover, the regulatory alternatives available entail particularly important efficiency losses. Imposing a uniform provider structure across insurers deprives them of an important possibility to match their products with the preferences of their clientele. It also becomes difficult for them to offer managed-care alternatives, which require differing degrees of pre-commitment by providers to accept rationing as a means to control moral hazard. The alternative of having insurers offer contracts featuring one type of provider only would make economics of scope impossible, as pointed out in the preceding paragraph. In all, a specialization of health insurers cannot occur.

Conclusions 4 through 6 together make it clear that not only present-day but just about any imaginable risk adjustment schemes cannot re-establish the absence of risk adjustment effort induced by uniform premiums, resulting in failure to correct for the deviation from static efficiency (see Figure 2).

## 4. Uniform premium regulation, risk adjustment, and innovation effort

During the past two decades, innovation effort by health insurers has revolved around similar issues as those discussed in the preceding section, viz. the level of benefits, structure of benefits, and structure of providers. In the following, the effect of uniform premium regulation on these three dimensions of innovation are studied, taking again into account risk adjustment as the typical secondary regulation complementing the imposition of uniform premiums (see Figure 2).

(a) **Innovation w.r.t. the level of benefits:** This amounts to offering contracts with different deductibles and rates of coinsurance as well as premium rebates

for no claims (which permit insured to determine their degree of cost sharing after the advent of illness). For instance, the Technicians' sickness fund, which belongs to German statutory health insurance, launched a contract with cost sharing in 2002 (www.tk-online.de), while Swiss social health insurers offer a choice of annual deductibles up to US$1,200 (at 2004 exchange rates). These efforts can be interpreted as attempts to adjust the properties of health insurance contracts to the presence of both *ex ante* and *ex post* moral hazard effects, expounded in Section 2. Since these alternatives expose the insured to differing degrees of financial risk, they must be priced differently. Full uniformity of premiums would doom these alternatives to failure to begin with.

Assuming that the regulator accepts some premium differentiation while continuing a risk adjustment scheme, the incentive for innovation is still undermined, and for the following reason. Those who choose to move to an innovative plan tend to be of younger age (see e.g. Cutler and Reber, 1998). However, all known risk adjustment schemes rely heavily on age as a predictor of HCE. As argued in item (1′) above in Conclusion 4, age adjustment is biased; specifically, it is likely excessive because left-out variables (such as previous HCE) correlate positively with age. An innovative insurer, showing an increased share of young enrolees on his books, therefore is subject to a financial sanction that exceeds the loss of revenue that would be induced by lower risk-based premiums for the young. In addition, such an insurer is not fully credited for a reduction of HCE achieved, to the extent that the schemes contain a retrospective element (which means that high HCE incurred are partially paid for and low HCE partially taxed). These two considerations imply that innovation with respect to the level of benefits occurs to a suboptimal degree. This is even true of an innovation that serves to reduce administrative cost because the young are again more likely to react to the concomitant reduction in premiums.

(b) Innovation w.r.t. the benefit structure: Here, an important innovation is the design of special disease management programs that coordinate the different benefits (in terms of different types of ambulatory and hospital care as well as drug treatments) to achieve the desired outcome. To the extent that the insured must give up freedom of choice in the course of a treatment path, they have to be offered a reduction in premium. However, disease management may also be seen as an investment into quality assurance by the insurer, justifying a higher premium. Again, this is at odds with full uniformity of premiums. The point is that uniform premiums within type of plan are not compatible with this type of innovation either.

One problem is that individuals opting for disease management programs are highly informed, which makes them particularly susceptible to moral hazard (see Section 3.3, item (1′) above Conclusion 5). This constitutes a left-out factor that cannot even be corrected for by diagnostic information, which by the way is needed for honoring innovative efforts in the guise of disease management programs. Furthermore, disease management may well reduce HCE by avoiding duplication of diagnostic tests and

treatments but increase administrative expense. However, existing risk adjustment schemes exclude administrative expense, a fact that discourages innovation of this type. Finally, within a given risk group, younger individuals are more likely to opt for such programs, as argued in the preceding paragraph. This means that innovative effort is again punished to some degree by a risk adjustment mechanism.

(c) Innovation w.r.t. the provider structure: An important innovation in this field is to contract with service providers who accept new forms of remuneration. In order to elicit sufficient effort for cost reduction and maintenance of quality, a two-part fee with some degree of supply-side cost sharing is optimal (Chalkley and Malcomson, 1998). Typically, younger physicians participate in such plans, which are usually associated with managed care and health maintenance organizations in particular. Since young physicians tend to have a younger clientele, an insurer launching an innovation involving a changed provider structure again attracts young individuals, which causes a financial sanction in all known risk adjustment schemes.

**Conclusion 7:** For all three types of innovation considered, insurer incentive to launch them is annihilated by fully uniform premiums and still attenuated by premiums differentiated by plan type combined with risk adjustment.

It may be worthwhile to recall that a basic motivation for introducing risk adjustment was to direct insurer effort away from risk selection and towards innovation efforts. However, the very same risk adjustment is now found to weaken consistently incentives to innovate. This is not compatible with dynamic efficiency.

## 5. Risk-based premiums and transfers as an alternative

Clearly, uniform premium regulation in health insurance causes important losses in terms of static (Conclusions 4 to 6) and dynamic efficiency (Conclusion 7). This motivates reconsideration of the first-best alternative. Given public information about risk, the first-best Pareto-optimum equilibrium with fair premiums is full coverage for high and low risks, associated with high and low premiums reflecting the difference in risk (Rothschild and Stiglitz, 1976). And, while the full coverage solution cannot be reached under conditions of administrative expense or moral hazard, it remains true that a set of contracts reflecting risk differences is Pareto optimal. Interestingly, risk selection activities by insurers now appear in a favourable light, because they reduce the information asymmetry that is the ultimate reason for these optima allegedly not being attainable.

Indeed, the difficulties encountered with imposing uniform or at least community-rated premiums and fully neutralizing risk selection incentives

through risk adjustment provide testimony to the effect that health insurers are to a considerable degree capable of gathering information about risk. In future, genetic information (if not prohibited) will go a long way towards the first-best equilibrium in health insurance markets. Letting insurers set premiums according to risk has the two following benefits.

- **Static efficiency:** Insurers can weigh the cost and expected benefits of information gathering about risk so as to arrive at an optimal degree of premium differentiation according to risk. To the extent that they indeed equalize the margin of premium over expected cost, the incentive to exclude risks of any type should be zero and the first-best Pareto optimum reached. However, recent evidence (Herring and Pauly, 2001) suggests that, even in the unregulated market for individual risks, US health insurers do not fully differentiate premiums according to risk. This points to the possibility of exclusion of some very high risks. A pool for these risks (assigned risk pool) may still be necessary.
- **Dynamic efficiency:** With risk-based premiums, consumers obtain the appropriate signals indicating to them that, by opting for contracts that induce more preventive effort or a commitment to limit *ex post* moral hazard, they can reap the benefit of future lower premiums. In turn, health insurers who develop new plans that attract individuals with high future HCE can count on receiving a higher premium. More generally, developing contracts that appeal to particular segments of the population does not meet with the regulator's suspicion that they are motivated by an interest in risk selection, but disguised as innovation. In addition, innovations entailing higher administrative expense but lowered HCE are not discriminated against by a risk adjustment scheme that pays for high HCE only.

The main concern with risk-based premiums is of course equity. Particularly in the context of health insurance, there are two conceptions of equity. One, adopted e.g. by Van de Ven and Ellis (2000), considers the fact that a high risk has to pay a high competitive premium as a violation of fairness, the argument being that individuals should not be punished for their unfavourable endowments. However, this implies that a rich person may be cross-subsidized simply because he or she is a high-risk type. The other conception, espoused by most economists, is that the health insurance premium to be paid may pose an excessive burden on the individual. According to this conception of equity, citizens characterized by the combination of high risk and low income and wealth must be enabled to buy (expensive) health insurance coverage.

A tax-financed, earmarked transfer achieves this objective. The proposed solution is to subsidize health insurance premiums to the extent that they exceed a certain percentage of household income (with the subsidy less than complete in order to safeguard incentives for search). Since competitive premiums reflect risk factors, this comes close to the risk-adjusted subsidies proposed by Pauly *et al.* (1992) and Van de Ven *et al.* (2000). In sum, health insurance is

relieved from the task of systematic redistribution. Competitive insurance is designed for redistribution governed by chance. Redistribution involving systematic differences between incomes or between health risks is the task of government, who has power to tax and to pay subsidies to achieve the equity goal.

Although of minor relevance for citizens, a major obstacle to such reform comes from the fact that the budgets of public health insurance and the government are not (fully) integrated. Politicians in government may resent having their budgets charged with subsidies that clearly go to the 'sick and poor'.

In order to assess the magnitude of budgetary restructuring, Zweifel and Breuer (2002) calibrated a loss density function with a mean that coincides with the average HCE of German statutory health insurers. Three draws from this distribution were averaged and assigned as the fair premium to the individuals of the official household survey. For simplicity no loading for administrative expense was added. The maximum share of total household income taken by health insurance is set at 15%, with the gap between this threshold and the calculated fair premium making up the subsidy. In this way, the total amount of subsidies can be calculated based on the joint income and HCE distribution. The estimated transfer volume (for 2000, at 2004 exchange rates) amounts to roughly $510 per capita or US$42bn in the aggregate.

However, in Germany employers pay part of the contributions to health insurance. If these are paid out to workers, becoming part of taxable labour income, tax revenue is estimated to increase by US$17bn. In this way, the net transfer to be financed out of the public budget is some US$18bn. In addition, unemployment insurance pays the health insurance premiums for the unemployed, and the public pension scheme, part of the premiums of retired persons. These payments amount to about US$20bn. They need not be made anymore once the proposed premium subsidies come directly from the public purse. Therefore, the net charge to the public purse due to the reform approaches zero. At the same time, the volume of transfers is reduced from its current value, occurring within German statutory health insurance, since transfers are more targeted.

Since mixed financial arrangements of this type (or tax reductions for health insurance payments made by employers) are widespread, one may state

**Conclusion 9:** A rough estimate of the transfer volume needed to complement risk-based health insurance premiums reveals that in the case of Germany, offsetting flows may be sizable enough as to result in a minimal net charge to the public purse. This result is expected to be typical for other industrial countries relying on tax-shielded employer's contributions to health insurance premiums.

## 6. Summary and conclusions

This contribution argues for risk-based premiums also in public health insurance. It reviews the properties of optimal insurance contracts in the presence of both *ex ante* and *ex post* moral hazard to conclude that uniformity of premiums is not compatible with optimality. On the part of insurers, the imposed uniformity of premiums creates an incentive for risk selection. Risk adjustment is designed to neutralize this incentive and redirect insurers' efforts towards innovation. It thus constitutes a secondary regulation induced by premium regulation and may be considered as part of a regulatory spiral that will result in a more and more refined (and costly) scheme. However, even such refinements (like diagnostic cost groups and pharmaceutical cost groups) are unlikely to neutralize fully the risk selection incentive; in addition, they punish innovative effort on the part of insurers. These efficiency losses can be avoided by letting insurers charge risk-based premiums, complemented by a tax-financed transfer mechanism. An earmarked subsidy that limits payments for health insurance to 15% of labour income before deductions was shown to affect the public budget only minimally, at least in the case of Germany. Therefore, a transition towards a more efficient but also equitable way of financing health insurance does not necessarily require a sizable restructuring of financial flows between social security and the government.

## References

Beck, K. and P. Zweifel (1998), 'Cream-skimming in deregulated social health insurance: evidence from Switzerland', in P. Zweifel (ed.), *Health, the Medical Profession, and Regulation*, Dordrecht: Kluwer, pp. 211–227.

Blomqvist, A.G. (1997), 'Optimal non-linear health insurance', *Journal of Health Economics*, **16**(3): 303–321.

Buchanan, J.L., E.B. Keeler, J.E. Rolph, and M.R. Holmer (1991), 'Simulating health expenditures under alternative insurance plans', *Management Science*, **37**(9): 1067–1089.

Chalkley, M. and J.M. Malcomson (1998), 'Contracting for health services with unmonitored quality', *Economic Journal*, **108**(449): 1093–1110.

Cutler, D.M. and S.J. Reber (1998), 'Paying for health insurance: the trade-off between competition and adverse selection', *Quarterly Journal of Economics*, **113**(May): 433–466.

Ehrlich, I. and G.S. Becker (1972), 'Market insurance, self-insurance, and self-protection', *Journal of Political Economy*, **69**: 355–368.

Enthoven, A.C. (1986), 'Managed competition in health care and the unfinished agenda', *Health Care Financing Review* (Annual Supplement), **7**: 105–120.

Glazer, J. and T.G. McGuire (2000), 'Optimal risk adjustment in markets with adverse selection: an application to managed care', *American Economic Review*, **90**: 1055–1071.

Glied, S. (2000), 'Managed care', in J.P. Newhouse and A.J. Culyer (eds), *Handbook of Health Economics*, Vol. 1A, Amsterdam: North-Holland, Chapter 13.

Gollier, C. (2000), 'Optimal insurance design: what can we do with and without expected utility?', in G. Dionne (ed.), *Handbook of Insurance*, Dordrecht: Kluwer, pp. 97–116.

Herring, B. and Pauly, M.V. (2001), 'Premium variation in the individual health insurance market', *International Journal of Health Care Finance and Economics*, **1**(1): 43–58.

Machina, M.J. (1995), 'Non-expected utility and the robustness of the classical insurance paradigm', *Geneva Papers on Risk and Insurance Theory*, **20**: 9–50.

Pauly, M.V. (1974), 'Overinsurance and public provision of insurance: the roles of moral hazard and adverse selection', *Quarterly Journal of Economics*, **88**(1): 44–62

Pauly, M.V. (1984), 'Is cream skimming a problem for the competitive medical market?', *Journal of Health Economics*, **3**: 87–95.

Pauly, M.V. *et al*. (1992), *Responsible National Health Insurance*, Washington, DC: AEI Press.

Rothschild, M. and J. Stiglitz (1976), 'Equilibrium in competitive insurance markets: an essay in the economics of incomplete information', *Quarterly Journal of Economics*, **90**: 629–649.

Van de Ven, W.P.M.M. and R.C.J.A. Van Vliet (1992), 'How can we prevent cream skimming in a competitive health insurance market?', in P. Zweifel and H.E. Frech (eds), *Health Economics Worldwide*, Dordrecht: Kluwer, pp. 23–46.

Van de Ven, W.P.M.M. and R.P. Ellis (2000), 'Risk adjustment in competitive health plan markets', in J.P. Newhouse and A.J. Culyer (eds), *Handbook of Health Economics*, Vol. 1A, Amsterdam: North-Holland, Chapter 14.

Van de Ven, W.P.M.M. *et al*. (2000), 'Access to coverage for high risks in a competitive individual health insurance market: via premium rate restrictions or risk-adjusted premium subsidies?', *Journal of Health Economics*, **19**: 311–339.

Van Vliet, R.C.J.A (1992), 'Predictability of individual health care expenditures', *Journal of Risk and Insurance*, **59**(3): 443–460.

Wilson, C.A. (1977), 'A model of insurance markets with incomplete information', *Journal of Economic Theory*, **16**(2): 167–207.

Zeckhauser, R. (1970), 'Medical insurance: a case study of the tradeoff between risk spreading and appropriate incentives', *Journal of Economic Theory*, **2**(1): 10–26.

Zweifel, P. and M. Breuer (2002), 'Weiterentwicklung des deutschen Gesundheitssystems. Gutachten im Auftrage des Verbandes Forschender Arzneimittelhersteller (Reform of the German health care system', Expert report for the Association of Research-based Pharmaceutical Companies), mimeo, University of Zurich, Zurich.

Zweifel, P. and F. Breyer (1997), *Health Economics*, New York and Oxford: Oxford University Press.

Zweifel, P. and W.G. Manning (2000), 'Moral hazard and consumer incentives in health care', in J.P. Newhouse and A.J. Culyer (eds), *Handbook of Health Economics*, Vol. 1A, Amsterdam: North-Holland, Chapter 8.